

# Automatic Conversion of English Pronunciation Using Sequence-to-Sequence Model

Kong Joo Lee<sup>†</sup> · Yong Seok Choi<sup>††</sup>

## ABSTRACT

As the same letter can be pronounced differently depending on word contexts, one should refer to a lexicon in order to pronounce a word correctly. Phonetic alphabets that lexicons adopt as well as pronunciations that lexicons describe for the same word can be different from lexicon to lexicon. In this paper, we use a sequence-to-sequence model that is widely used in deep learning research area in order to convert automatically from one pronunciation to another. The 12 seq2seq models are implemented based on pronunciation training data collected from 4 different lexicons. The exact accuracy of the models ranges from 74.5% to 89.6%. The aim of this study is the following two things. One is to comprehend a property of phonetic alphabets and pronunciations used in various lexicons. The other is to understand characteristics of seq2seq models by analyzing an error.

**Keywords :** Phonetic Alphabet, Word's Pronunciation, Allophone of English Word, Sequence-to-Sequence Model

## Sequence-to-Sequence Model을 이용한 영어 발음 기호 자동 변환

이 공 주<sup>†</sup> · 최 용 석<sup>††</sup>

## 요 약

영어는 동일 철자의 발음이 매우 다양한 언어이기 때문에 사전에 기술되어 있는 단어의 발음기호를 읽어야만 정확한 발음을 알 수 있다. 영어 사전마다 사용하는 발음기호(phonetic alphabet) 시스템이 다르며 같은 단어에 대해 기술하고 있는 발음 역시 다르다. 본 연구에서는 최근 딥 러닝 분야에서 널리 사용되고 있는 sequence-to-sequence (seq2seq) model을 이용하여 사전마다 다른 발음을 자동으로 변환해 보고자 한다. 4가지 다른 종류의 사전에서 추출한 발음 데이터를 이용하여 모두 12개의 seq2seq model을 구현하였으며, 발음 자동 변환 모듈의 정확 일치율은 74.5% ~ 89.6%의 성능을 보였다. 본 연구의 주요 목적은 다음의 두 가지이다. 첫째 영어 발음기호 시스템과 각 사전의 발음 데이터 특성을 살펴보는 것이고, 둘째, 발음 정보의 자동 변환과 오류 분석을 통해 seq2seq model의 특성을 살펴보는 것이다.

**키워드 :** 발음기호, 단어 발음, 영어 단어 발음의 다양성, Sequence-to-Sequence Model

## 1. 서 론

영어는 동일한 음소(phoneme)가 서로 다른 철자로 기술될 수 있는 언어이다[1]. 그렇기 때문에 영어 단어의 철자만으로는 정확한 발음을 예측해 내기 어렵다. 영어 단어의 정확한 발음을 알기 위해서는 사전에 기술되어 있는 발음기호를 읽어야 한다. 또한, 음성인식이나 음성합성 등과 같이 음성을 다루는 응용 시스템 구축에는 각 단어의 발음기호에

대한 이해가 필수적이다.

공개되어 있는 영어 사전들이 사용하는 발음기호는 사전마다 다르다. 가장 많이 사용하는 발음기호 시스템은 IPA (International Pronunciation Alphabet)이다[1]. Table 1은 단어 'gentle'과 'jurisdiction'의 발음 정보를 여러 사전에서 사용하는 다양한 발음기호 시스템으로 기술해 놓은 것이다.

Table 1의 'ARPAbet'은 미국 영어의 발음 정보를 기계적으로 처리하기 위하여 제안된 발음기호이며, 그렇기 때문에 ASCII 심볼만으로 구성되어 있다[1].

'MWCD'는 Merriam-Webster's Collegiate Dictionary에서 쓰이는 발음기호를 의미하며, 'NOAD'는 New Oxford American Dictionary에서 쓰이는 발음기호이다.

※ 이 연구는 충남대학교 학술연구비에 의해 지원되었음.  
† 종신회원 : 충남대학교 전자정보통신공학과 교수  
†† 준 회 원 : 충남대학교 전자전파정보통신공학과 석사과정  
Manuscript Received : January 18, 2017  
Accepted : February 10, 2017  
\* Corresponding Author : Kong Joo Lee(kjoollee@cnu.ac.kr)

Table 1. Various Phonetic Alphabets for Transcribing Pronunciations

	gentle	jurisdiction
IPA	'dʒɛntəl	.dʒʊəɾɪs'dɪkʃ(ə)n
ARPAbet	JH EH1 N T AH0 L	JH UH2 R AH0 S D IH1 K SH AH0 N
MWCD	'jen-t'l	jʊr-əs-'dik-shən
NOAD	'jen(t)l	.jʊərəs'dikSH(ə)n

ARPAbet은 각 음소마다 띄어쓰기로 구분을 해 놓았고 MWCD의 경우에는 각 음절(syllable)마다 '-'로 구분을 해 놓았다. Table 1에서 보는 바와 같이 사전마다 사용하는 발음기호가 다르기 때문에 영어 학습자에게는 부담으로 작용할 수 있다. 또한 음성처리를 위한 응용 시스템 구축에서도 단어의 발음 정보를 수집할 때 서로 다른 발음기호의 사용은 사전 사용의 제한을 가져올 수 있다.

사전마다 사용하는 다른 발음기호는 일대일의 단순한 매핑으로 변환이 쉽지 않다. Table 2의 예를 보면 NODE (New Oxford Dictionary of English) 사전에서 단어 'gem'의 /ɛ/는 CMUdict(Carnegie Mellon University Pronouncing Dictionary)와 MWCD(Merriam-Webster Collegiate Dictionary)에서는 /EH/ 또는 /e/로 매핑되는 반면, 단어 'congress'의 경우에는 /AH/ 또는 /ə/로 매핑되었다.

Table 2. Examples of Various Mapping from Phoneme /ɛ/

	gem	congress
NODE (IPA)	dʒɛm	'kɒŋgrɛs
CMUdict (ARPAbet)	JH EH1 M	K AA1 NG G R AH0 S
MWCD (MWCD)	'jem	'kæŋ-græs

이와 같은 이유로 서로 다른 발음기호를 사용하는 사전 쌍에서 하나의 발음기호에서 다른 발음기호로 일대일 매핑을 통한 단순 변환이 가능하지 않다. 또한 같은 발음기호 시스템을 사용한다 하더라도 사전이 주로 영국 영어를 다루는지 미국 영어를 다루는지에 따라 기술된 발음에 차이가 있다.

영국 영어와 미국 영어의 발음 차이[2] 중, 본 논문에서는 여러 단어에 걸쳐 나타나면서 사전의 발음기호에서 그 차이를 구분할 수 있는 대표적인 현상만을 다루었다.

Table 3은 영국 영어와 미국 영어에서 자음 발음의 대표적인 차이를 보인다. 미국 영어에서는 한 음절 내에서 모음 뒤에 나오는 /r/을 발음하는 반면, 영국 영어에서는 모음 앞

을 제외하고는 대부분의 경우 /r/을 발음하지 않는다. 또한, 미국 영어에서 /t/는 모음 사이에서 /r/로 약화되거나 종종 생략되기도 한다.

Table 3. Consonant Allophone Different from British English and American English

	rhotic consonant /r/	realization of /t/
British English	/r/	/t/
American English	/r/	/r/ or /(t)/

Table 4. Examples of Consonant Allophone in Various Kinds of Lexicons

	linger	hard	bottle	winter
NODE (IPA) (British)	'lɪŋgə	hɑ : d	'bɒt(ə)l	'wɪntə
NOAD (NOAD) (American)	'liNGgər	hɑrd	'bɑdl	'wɪntər 'wɪnər

Table 4에서는 영국 영어를 다루고 있는 사전 NODE와 미국 영어를 다루는 NOAD 사전에서 추출한 발음 정보를 비교하였다. NODE 사전은 'linger', 'hard' 단어의 경우 Table 3에서 설명한 바와 같이 음소 /r/이 생략된 것을 확인할 수 있다. 또한 단어 'bottle'은 미국 영어인 NOAD에서는 't'가 /d/로 표기되었고, 'winter'의 경우에는 't'가 생략된 발음 /wɪnər/가 포함되어 있는 것을 확인할 수 있다.

Table 5는 모음에서 나타나는 영국 영어와 미국 영어의 대표적인 차이를 제시한다.

본 연구에서는 서로 다른 사전에서 추출한 발음 정보를 자동으로 변환시켜주는 모듈을 음소 단위의 sequence-to-sequence(seq2seq) model을 이용하여 구축해 보고자 한다. Seq2seq model은 Recurrent Neural Network (RNN)의 일종으로 가변 길이의 입력을 하나의 벡터로 표현한 후, 이로부터 가변 길이의 출력을 생성할 수 있는 신경망이다[5]. 이를 위하여 서로 다른 발음기호 시스템을 사용하는 온라인 사전으로부터 단어와 그 발음 정보를 수집하고 이를 학습데이터로 사용하여 seq2seq model을 학습시킨다. 실제 구현은 tensorflow에서 제공하는 seq2seq model[4]을 기반으로 구축하였다.

본 논문에서는 이를 통해 두 가지를 얻고자 한다. 첫째, 사전에 수록되어 있는 발음 정보들 사이를 자동으로 변환시킬 수 있는 모듈을 구축하고, 그 결과 영어 발음 정보에 대한 특성을 살펴본다. 둘째, 작은 데이터 집합을 이용하여 seq2seq model의 특성을 분석해 본다.

1) 각 사전과 사전에서 사용하는 발음기호에 대한 설명은 3.1절에서 자세히 다룬다.

Table 5. Vowel Allophone Different from British and American English and Its Examples

	want		class		promotion	
Naver (IPA) (British)	wɒnt	/ɒ/	klɑ:s	/ɑ : /	prəlməʊʃn	/əʊ/
Naver (IPA) (American)	wɑ:nt wɔ:nt	/ɑ : / /ɔ : /	klæs	/æ/	prəlməʊʃn	/oʊ/

## 2. 관련 연구

### 2.1 음차 표기

음차 표기는 한 언어의 문자 체계를 다른 문자 체계로 나타내는 것이다. 기계번역, 정보 검색 등에서 단어 불일치 문제를 해결하기 위해 다양한 자동 음차 표기 연구가 진행되고 있다. [6]에서는 구문기반의 통계적 기계번역(PBSMT) 방법을 활용하여 음차 표기를 자동으로 변환하는 시스템을 연구하였다. [7]은 [6]의 기반에 RNN 언어모델을 활용하여 출력 결과의 순위를 재구성하였다. [8]에서는 sequence-to-sequence 모델을 도입하였는데, seq-to-seq 모델만으로 구성된 음차표기 시스템보다는 PBSMT의 결과 순위를 seq-to-seq 모델을 이용하여 재구성한 시스템이 9.7%의 성능 향상을 보였다.

### 2.2 Seq2seq Model

기계번역은 하나의 입력 시퀀스를 다른 출력 시퀀스로 변환하는 대표적인 문제이다[3]. 입력 시퀀스와 출력 시퀀스는 그 길이가 매우 다양할 수 있다. 그렇기 때문에 입력 시퀀스 길이의 다양성을 효율적으로 표현하는 것이 매우 중요하다. RNN은 다양한 길이의 입력을 고정된 크기의 벡터값으로 매핑시키는데 효과적인 구조를 갖고 있다.

최근에는 이와 같은 sequence-to-sequence 문제를 딥 러닝 알고리즘을 이용하여 좋은 연구 결과를 얻고 있다. 일반적으로 sequence-to-sequence 문제는 인코더로 사용되는 RNN와 디코더로 사용되는 RNN을 연결하여 구현한다. 인코더가 입력 시퀀스를 고정된 크기의 벡터값으로 매핑하면, 디코더는 이 벡터값으로부터 다시 다양한 길이의 시퀀스를 생성해 낸다[3].

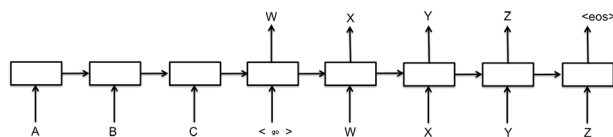


Fig. 1. A Basic Architecture of Sequence-to-Sequence Model[5]

Fig. 1은 일반적인 sequence-to-sequence 모델의 처리 방법을 개략적으로 보여주고 있다. sequence-to-sequence 모델은 입력 시퀀스 'A B C'을 하나씩 읽어들이어서 고정된 크기의 벡터를 만들어 낸다. 고정된 크기의 벡터로부터 다시

새로운 시퀀스 'W X Y Z'를 <EOS>가 출력될 때까지 만들어 낸다.

본 논문에서는 TensorFlow에서 제공하는 seq2seq model을 이용하여 발음의 자동 변환을 구현하였다. 또한, TensorFlow에서 제공하는 RNN의 제약사항 중 하나인 static RNN의 사용으로 인해 입력력 시퀀스의 길이에 따른 버킷팅을 도입하였다. 이는 시퀀스의 입력력 길이의 변화에 따른 불필요한 패딩을 최소화하고 효율적인 RNN 학습을 위하여 미리 정의된 버킷별로 학습데이터를 재분류하는 작업이다[4]. TensorFlow의 RNN은 long-distance dependency 문제를 완화하기 위하여 입력 시퀀스를 거꾸로 입력하여 학습시킨다. Fig. 1의 예제의 경우 입력과 출력의 버킷 크기 <5, 10>일 때, 입력은 [PAD, PAD, C, B, A]가 되며 출력은 [GO, W, X, Y, Z, PAD, PAD, PAD, PAD, PAD]이 된다.

## 3. 발음 사전 추출

### 3.1 사전 구성 및 특성

본 연구에서 사용한 사전 정보는 Table 6과 같다.

각 사전의 출처는 다음과 같다.

CMUdict: <http://www.speech.cs.cmu.edu/cgi-bin/cmu dict>

NOAD: <https://en.oxforddictionaries.com/definition/us/>

MWCD : <https://www.merriam-webster.com/>

NODE: <https://en.oxforddictionaries.com/>

각 사전에 대한 특징을 간략히 정리해 보면 다음과 같다.

1) **CMUdict** (Carnegie Mellon University Pronouncing Dictionary)는 약 134,000 단어에 대한 북미 영어 발음 사전으로 주로 음성 인식기 개발에 사용할 목적으로 구축되었다. 발음기호는 **ARPAbet**을 사용하여 기술하였으며, 일반 사전과 달리 'flowers', 'makes', 'picked'와 같이 활용형태의 단어들도 다수 포함되어 있다.

2) **NOAD** (New Oxford American Dictionary)는 현대 미국식 영어를 기반으로 구축되어 있기 때문에 미국 영어 발음의 특성이 다수 포함되어 있다. NOAD 자체의 발음기호를 사용한다.

3) **MWCD** (Merriam-Webster Collegiate Dictionary)는 미국 영어를 다루고 있는 사전이다. 다른 사전에 비해 단어별 발음 기호가 많이 포함되어 있다. (물론, 본 연구에서 수

Table 6. Summarization of Lexicons Used in This Paper

Phonetic Alphabet	CMUdict ARPAbet	NOAD NOAD	MWCD MWCD	NODE IPA
number of phonemes (consonant/vowel)	39 (23/16)	42 (24/18)	45 (25/20)	49 (26/23)
avg. number of phonemes per pronunciation	7.625	8.2514	8.413	8.030
avg. number of stresses per pronunciation	1.2596	1.24096	1.3768	0.94009
avg. number of pronunciations per entry	1.069	1.293	1.925	1.313
total number of entries	117,413	24,727	24,988	24,458

집한 발음에는 어린이용, 의학용, 영어학습자용 단어의 발음 기호는 포함시키지 않았다.) Merriam-Webster사는 자체 발음기호를 사용하여 단어의 발음을 표시한다.

4) **NODE** (New Oxford Dictionary of English) 사전은 주로 영국 발음의 특징을 많이 포함하고 있다. 발음기호는 **IPA**를 따른다.

3.2 사전 추출 및 전처리

온라인 사전으로부터 seq2seq model에서 사용할 발음기호를 추출하는 방법은 다음과 같다.

- (1) 온라인 사전에서 각 엔트리 별로 발음기호를 추출한다.
- (2) 부분 발음 정보를 확장하여 완성한다.

온라인 사전의 발음 정보는 주로 사전을 검색하는 사람을 위해 기술되어 있기 때문에 여러 개의 발음이 가능할 경우, 동일한 발음 부분이 최대한 반복되지 않도록 기술하였다. 그

래서 사람에게는 이해가 쉬우나 기계가 처리하기에는 다소 어려운 부분이 있다. 본 연구에서는 전처리로 이와 같은 반복 부분을 모두 확장하여 완성된 발음 정보를 추출하였다.

Table 7의 예제는 MWCD에 기술되어 있는 발음 정보의 일부이다. 예제 (a)은 ‘acceleration’의 발음이다. ‘original pronunciation’ 열에는 표면적으로는 2개의 발음이 기술되어 있으나 2번째 발음은 공통되는 부분을 제외하고 첫 번째 발음과 달라지는 부분만을 기술하였다. 또한, 두 번째 발음의 강세는 괄호 안에 포함되어 있어 선택적임을 알 수 있다. 이를 모두 확장하여 완성한 발음을 ‘extended pronunciation’ 열에 나타내었다.

예제 (c)과 (d)의 경우에는 두 개의 발음 정보가 기술되어 있는데, 두 번째 발음은 완성된 최종 발음에 포함되지 않았다. 첫 번째 발음을 처리하고 나았을 때, 음절 수가 다른 발음이 만들어지면 이를 기반으로 완성시켜야 하는 부분 발음의 확장을 수행하지 않았다.

Table 7. Examples of Pronunciations Described in MWCD

	entry	original pronunciation	extended pronunciation
(a)	acceleration	/ɪk-,se-lə-ˈrɑ-shən, (,)ak-/	ik-,se-lə-ˈrɑ-shən ak-,se-lə-ˈrɑ-shən ak-,se-lə-ˈrɑ-shən
(b)	accurately	/ˈɑ-kyə-rət-lē, ˈɑ-k(ə)rət-, ˈɑ-k(y)ərt-/	ˈɑ-kə-rət-lē ˈɑ-kyərt-rət-lē ˈɑ-kərt-rət-lē ˈɑ-krət-rət-lē ˈɑ-kyə-rət-lē
(c)	proselytization	/ˌprɑ-s(ə)lə-tə-ˈzɑ-shən, ˌprɑ-sə-ˌli-tə-/	ˌprɑ-slə-tə-ˈzɑ-shən ˌprɑ-sə-lə-tə-ˈzɑ-shən
(d)	conference	/ˈkɑn-f(ə)rən(t)s, -fərn(t)s/	ˈkɑn-fə-rənts ˈkɑn-fə-rəns ˈkɑn-frənts ˈkɑn-frəns
(e)	debris	/də-ˈbrē, dā-, ˈdā-, ˈde-(,)brē/	də-ˈbrē dā-ˈbrē    ˈdā-,brē ˈde-,brē    ˈde-brē

(3) 강세(stress) 정보를 처리한다.

영어에서 강세 정보는 단어의 품사나 의미를 변화시킬 수 있을 만큼 중요하다. 그러나 각 사전마다 강세를 기술하는 방법뿐 아니라 동일한 단어에 포함시킨 강세의 개수가 다른 경우도 허다하다. Table 8은 각 발음기호 시스템마다 사용하는 강세 기술 방법이다. IPA/MWCD/NOAD의 경우에는 음절 앞쪽에 강세를 위치시키는 반면, ARPAbet의 경우에는 해당 음소의 뒤에 0, 1, 2 숫자를 붙여 강세를 표시한다.

Table 8. Description of Stress According to Phonetic Alphabet

Phonetic Alphabet	primary stress	secondary stress	no stress
IPA/MWCD/NOAD	[ˈ]	[ˌ]	
ARPAbet	1	2	0

Table 9는 각 사전에 포함되어 있는 강세의 예제를 보여 준다. 동일한 단어에 대해 각 사전마다 포함하고 있는 강세 정보가 동일하지 않음을 알 수 있다.

Table 9. Different Usage of Stresses According to Lexicons

	insignia	rare	rehab
CMUdict	IH2 N S IH1 G N IY2 AH0	R EH1 R	R IY0 HH AE0 B
NOAD	in'signēə	rer	'rē,hab
MWCD	in-'sig-nē-ə	'rer	'rē-hab
NODE	ɪn'sɪɡnɪə	rɛ :	'ri : hab

본 연구에서는 발음 정보 간의 자동변환을 학습시켜야 한다. 그렇기 때문에 각 사전에서 강세를 표시하는 방법을 통일시켰다. 강세를 해당 음소의 바로 뒤쪽에 독립된 음소처럼 표시하였다. 예를 들어, 'insignia'의 경우, /IH-%-N-S-IH-\*-%-G-N-IY-%-AH0/와 같이, 제1 강세(primary stress)는 /\*/로 제2 강세(secondary stress)는 /%/로 독립된 음소처럼 기술하고 강세를 받는 음절의 바로 다음에 독립된 음소처럼 표기하였다. 'rare'의 MWCD 사건의 경우, /R-E1-\*-%-R/와 같이 기술한다.

(4) non-ASCII code를 ASCII code로 매핑한다.

발음기호 ARPAbet을 제외하고는 발음기호 시스템에서 사용하는 음소는 non-ASCII code인 경우가 많다. 내부 처리의 편리성을 위해 non-ASCII code인 발음기호를 ASCII code로 매핑하였다. 또한, 각 음소를 쉽게 구분하기 위하여 음소 사이를 '-'로 나누었다. 각 사전에 대한 최종 처리 결과의 일부 예제를 Table 10에 제시하였다.

Table 10. Examples of Normalized Pronunciations for Inputs of seq2seq Model

	insignia	rare	rehab
CMUdict	IH-%-N-S-IH-*-%-G-N-IY-%-AH	R-EH-*-%-R	R-IY-HH-AE-B
NOAD	I1-N-S-I1-*-%-G-N-E2-E3	R-E1-R	R-E2-*-%-H-A1-%-B
MWCD	I1-N-S-I1-*-%-G-N-E2-E3	R-E1-*-%-R	R-E2-*-%-H-A1-%-B
NODE	I2-N-S-I2-*-%-G-N-I2-E3	R-E2:	R-I1:-*-%-H-A4-B

### 4. 실험 및 평가

#### 4.1 학습데이터 및 평가데이터

학습데이터 구성은 Table 11과 같다.

평가데이터의 경우에는 4개의 사전에 공통으로 발생한 단어 중, 1,000개를 선정해서 평가데이터로 구축하였다.

각 모델은 모델이 생성한 발음기호가 정답과 얼마나 정확히 일치하는가로 평가한다.

Table 11. Number of Pronunciations in Training Data

Training Data	number of pronunciations
MWCD(MWCD) ↔ NODE(IPA)	44,086
MWCD(MWCD) ↔ CMUdict(ARPAbet)	37,471
MWCD(MWCD) ↔ NOAD(NOAD)	45,549
CMUdict(ARPAbet) ↔ NODE(IPA)	34,664
CMUdict(ARPAbet) ↔ NOAD(NOAD)	35,082
NODE(IPA) ↔ NOAD(NOAD)	44,403

#### 4.2 seq2seq model 학습 환경

본 연구에서 다루는 사전은 4개이기 때문에 서로 다른 사건의 발음기호로 변환하기 위해서는 모두 12개의 모델을 구축해야 한다. 모델의 학습에 필요한 파라미터들을 각 모델마다 조절해야 하지만 본 연구에서는 두 개의 모델에 대해 조절한 파라미터를 다른 모델에도 그대로 적용하였다.

기본 실험을 통해서 얻은 모델의 정확률 중, 가장 좋은 성능을 보이는 2개의 모델에 대하여 다음과 같은 성능 평가를 통해 학습 파라미터 값을 결정하였다. (Table 12에서 15까지는 두 모델 성능의 평균값이다.)

Table 12는 seq2seq model의 EPOCH값을 40으로 고정하고 learning rate를 변화시켜 가며 수행한 결과의 정확률이다. Table 13은 learning rate을 0.1, 0.01과 0.05로 고정시키고 EPOCH수를 늘려가며 수행한 결과이다.

Table 14는 seq2seq 모델의 layer 수에 따른 정확률이다. layer 수가 2일 때 가장 좋은 성능을 보였다.



Table 12. Accuracy According to the Parameter 'learning rate' with 40 EPOCHs

learning rate	Accuracy
0.01	82.9
<b>0.05</b>	<b>87.3</b>
0.1	86.4
0.3	84.5
0.5	84.9
0.7	78.4

4.3 Seq2seq model의 버킷 크기에 따른 성능 비교

2장에서 언급한 바와 같이 seq2seq model은 입력 길이의 차이로 인한 문제를 최소화하기 위하여 버킷팅을 사용한다. Table 15는 seq2seq model 학습 시 버킷 크기에 따른 성능을 살펴보았다. 학습에 사용한 원시사건의 발음 길이는 3~16개 음소의 범위를 가진다. 평가데이터 1,000개 중, 507개의 경우에는 원시사건과 목적사전 발음의 길이가 동일했다. 버킷 개수가 5일 때 가장 좋은 성능을 보였다.

4.4 Seq2Seq model을 이용한 자동 발음 변환 성능

4.2절과 4.3절의 결과로부터 최종 학습에 사용한 파라미터는 다음과 같다. Seq2seq model의 hidden layer unit 개수는 50, 모델의 layer는 2를 사용하였고 버킷 개수는 5를 사용하였다. 모델에 따라 learning rate와 EPOCH수는 0.01, 160번,

또는 0.05, 50번을 사용하였다. Batch 크기는 32, learning rate decay는 0.99를 사용하였다.

Table 16은 실험 결과 각 모델의 정확률이다. NOAD (NOAD) → MWCD(MWCD) 변환이 가장 높은 성능을 보였다. 반면, NODE(IPA) → CMUdict(ARPAbet) 변환이 가장 낮은 성능을 보였다.

Table 17은 사전 쌍마다 양방향 성능의 평균값을 비교한 결과이다. NOAD(NOAD) ↔ MWCD(MWCD) 변환이 가장 좋은 성능을 보였다. 두 사전은 모두 미국 영어 발음을 기술하고 있어서 가장 유사도가 높은 사전이었다. 반면에 NODE(IPA) ↔ CMUdict(ARPAbet) 변환 결과가 가장 낮은 성능을 보였는데, 이는 영국 영어를 다루는 NODE가 포함된 결과가 하위 3위를 모두 차지하는 것을 보았을 때, 미국 영어와 영국 영어의 차이 때문인 것으로 판단된다.

Table 18은 Table 16을 변환의 원시사전과 목적사전별로 정확도 평균을 낸 것이다. 특이한 점은 원시사전으로 CMUdict를 사용할 경우 가장 좋은 성능을 발휘한 반면, 목적사전으로의 CMUdict는 가장 낮은 성능을 보였다. CMUdict는 4개 사전 중, 가장 적은 음소를 사용하는 ARPAbet으로 기술되어 있으며, 발음 당 사용하는 평균 음소 개수도 최소이다. 또한 목적사전별 성능 순은 목적사전의 발음 당 평균 음소 개수 순과 일치한다. 짧은 시퀀스로부터 긴 시퀀스를 생성해 내는 문제가 그 반대의 경우보다 좀 더 용이할 것이라고 추정할 수 있다.

Table 13. Accuracy According to Parameters 'learning rate' and 'EPOCH'

learning rate / EPOCH	Accuracy	learning rate / EPOCH	Accuracy	learning rate / EPOCH	Accuracy
0.1 / 20	80.2	0.05 / 40	87.3	0.01 / 40	82.9
0.1 / 30	86.1	0.05 / 50	<b>87.4</b>	0.01 / 60	85.1
0.1 / 40	<b>86.4</b>	0.05 / 60	85.9	0.01 / 80	86.9
0.1 / 50	86.0	0.05 / 70	87.3	0.01 / 100	86.9
0.1 / 60	86.3	0.05 / 80	86.3	0.01 / 120	87.1
0.1 / 70	86.2	0.05 / 90	86.4	0.01 / 160	<b>87.9</b>
		0.05 / 100	85.5	0.01 / 180	87.4
				0.01 / 200	86.5

Table 14. Accuracy According to the Number of Layers of seq2seq Model

number of layers	Accuracy
1	86.1
<b>2</b>	<b>87.9</b>
3	86.9
4	83.3

Table 15. Accuracy According to the Number of Buckets in seq2seq Model

number of buckets	Accuracy
Bucket-2	88.4
Bucket-3	84.5
Bucket-4	87.9
<b>Bucket-5</b>	<b>89.2</b>
Bucket-6	89.1
Bucket-7	88.7
Bucket-8	89.1
Bucket-9	88.1

Table 16. Accuracy According to the Models

	Model	Accuracy	learning rate / EPOCH
1	MWCD(MWCD) → NODE(IPA)	81.3	0.01/160
2	MWCD(MWCD) → CMUdict(ARPAbet)	82.0	0.01/160
3	MWCD(MWCD) → NOAD(NOAD)	87.8	0.01/160
4	CMUdict(ARPAbet) → NODE(IPA)	79.7	0.01/160
5	CMUdict(ARPAbet) → MWCD(MWCD)	88.9	0.01/160
6	CMUdict(ARPAbet) → NOAD(NOAD)	87.6	0.01/160
7	NODE(IPA) → NOAD(NOAD)	82.1	0.05/50
8	NODE(IPA) → CMUdict(ARPAbet)	74.5	0.05/50
9	NODE(IPA) → MWCD(MWCD)	83.8	0.01/160
10	NOAD(NOAD) → NODE(IPA)	77.9	0.05/50
11	NOAD(NOAD) → MWCD(MWCD)	<b>89.6</b>	0.05/50
12	NOAD(NOAD) → CMUdict(ARPAbet)	81.5	0.01/160

Table 17. Comparison of Average Accuracies of Bi-directional Models

Bi-direction model	Accuracy
NOAD(NOAD) ↔ MWCD(MWCD)	<b>88.7</b>
CMUdict(ARPAbet) ↔ MWCD(MWCD)	85.5
CMUdict(ARPAbet) ↔ NOAD(NOAD)	84.6
NODE(IPA) ↔ MWCD(MWCD)	82.6
NODE(IPA) ↔ NOAD(NOAD)	80.0
NODE(IPA) ↔ CMUdict(ARPAbet)	77.1

Table 18. Comparison of Average Accuracies According to the Same Source and Target Lexicons.

	Model	Accuracy
from	CMUdict(ARPAbet) →	85.4
	MWCD(MWCD) →	83.7
	NOAD(NOAD) →	83.0
	NODE(IPA) →	80.1
to	→ MWCD(MWCD)	87.4
	→ NOAD(NOAD)	85.3
	→ NODE(IPA)	79.6
	→ CMUdict(ARPAbet)	79.3

NODE(IPA)는 원시사전과 목적사전 모두에서 낮은 성능을 보였는데, NODE(IPA)가 다른 사전과 달리 영국 영어의 특성을 담고 있기 때문으로 파악된다.

#### 4.5 오류 결과 분석

본 절에서는 각 사전 변환에 대해 오류를 분석해 본다. 가장 많은 부분을 차지하는 오류에 대해서 다음과 같이 분류하였다.

**강세 오류**는 변환된 발음과 정답 발음 사이에 강세만이 차이가 나는 오류를 의미한다.

**모음교체 오류**는 변환된 발음과 정답 발음 사이에 사용된 모음이 서로 다른 경우를 의미한다.

**반복 오류**는 주로 단어 마지막에 동일한 음소가 반복해서 나타나는 오류를 의미한다.

**자음 오류**는 자음 하나가 교체되었거나, 삽입 또는 삭제된 경우를 의미한다.

Table 19는 가장 좋은 성능과 가장 낮은 성능을 보인 변환에 대해 그 오류를 분석한 결과이다. '1 edit-dist 오류'는 정답 발음과 모델이 생성한 발음을 편집 거리(edit distance)

Table 19. Error Analysis of Pronunciations Generated from the Best and the Worst Models

		NOAD(NOAD) → MWCD(MWCD)	NODE(IPA) → CMUdict(ARPAbet)
edit-dist length	error with 1 edit-dist	79	176
	error with 2 edit-dist	17	56
	error with 3 or more	8	23
	<b>total number of errors</b>	<b>104</b>	<b>255</b>
edit kind	insertion error <sup>2)</sup>	5	43
	deletion error	9	14
	replacement error	75	169
type of error	stress error <sup>3)</sup>	19 (18.3%)	48 (18.8%)
	vowel replacement error	<b>37 (35.6%)</b>	<b>118 (46.3%)</b>
	repetitive error	0	<b>14 (5.5%)</b>
	consonant-related error	<b>30 (28.8%)</b>	8 (2.0%)

2) 이 표에서 삽입/삭제/교체 오류에는 강세 오류는 제외되었다. 삽입/삭제/교체 오류는 중복해서 계산하였음.

3) 이 표에서 오류 종류별 분석은 단순 오류(한 종류의 오류만 있는 경우)에 대해서만 수행하였다.

알고리즘으로 비교했을 때, 편집 거리가 1인 오류를 의미한다. 또한, 편집 종류에 따라 삽입/삭제/교체 오류로 분류하였다. 표에서 보는 바와 같이, 두 경우 모두 '1 edit-dist 오류'가 제일 많은 부분을 차지하였고 편집 종류 중에는 교체 오류가 제일 많은 부분을 차지하였다.

(1) NOAD(NOAD) → MWCD(MWCD) 변환

모음교체 오류 37개 중 20개가 음소 /ə/에 대한 잘못된 매핑으로 발생하였다. Table 20은 모음교체 오류의 예제이다.

NOAD의 음소 /ə/는 MWCD에서 두 개의 서로 다른 음소로 매핑이 가능하다. 이 매핑을 모델이 학습하면서 잘못된 결과를 생성한 경우가 모음교체 오류의 대부분을 차지하였다.

Table 20. Examples of Vowel-replacement Errors in NOAD(NOAD) → MWCD(MWCD)

entry	NOAD	MWCD (truth)	MWCD (model)	description
rusk	/rʌsk/	/rʌsk/	/risk/	/ə/ vs. /i/ replacement
elate	/ə'lat/	/i-'lat/	/ə-'lat/	/ə/ vs. /i/
modest	/mädəst/	/mä-dəst/	/mä-dist/	/ə/ vs. /i/
seclude	/sə'klüd/	/si-'klüd/	/sə-'klüd/	/ə/ vs. /i/
demean	/də'men/	/di-'men/	/di-'men/	positive
jump	/jəmp/	/jəmp/	/jəmp/	positive

또한, NOAD(NOAD) → MWCD(MWCD) 변환에서 특이한 부분은 자음오류가 28% 이상 차지하고 있는 것이다. Table 4에서 보인 바와 같이 NOAD는 글자 't'의 발음이 /d/로 약화되는 경우가 사전에 포함되어 있다. NOAD(NOAD) → MWCD(MWCD) 학습데이터에서 음소 /d/를 그대로 /d/로 학습한 경우와 /t/로 변환 학습한 경우가 거의 동수로 나타났다. Table 19의 자음오류 30개 중, 24개는 NOAD의 /d/

Table 21. Examples of Consonant-replacement Errors in NOAD(NOAD) → MWCD(MWCD)

entry	NOAD	MWCD (truth)	MWCD (model)	description
paddy	/padē/	/pa-dē/	/pa-tē/	/d/ vs. /t/ replacement
cotta	/kädə/	/kä-tə/	/kä-də/	/d/ vs. /t/
comedy	/kämädē/	/kä-mə-dē/	/kä-mə-tē/	/d/ vs. /t/
sweater	/swedər/	/swe-tər/	/swe-dər/	/d/ vs. /t/
songwriter	/sɔNGrɪdər/	/sɔŋ-rɪ-tər/	/sɔŋ-rɪ-tər/	positive
girder	/gərdər/	/gər-dər/	/gər-dər/	positive

→/t/ 발음 상의 변화를 학습한 것이 잘못 적용되면서 발생한 오류였다. Table 21에 자음교체 오류의 예제를 보인다.

(2) NODE(IPA) → CMUdict(ARPAbet) 변환

이 변환은 전체 변환에서 가장 낮은 성능을 보인 경우이다. 여기서 가장 두드러진 것은 반복오류이다. 대부분의 반복오류가 발음 제일 마지막에 /NG/ 음소가 반복되는 경우였다. Table 22에 오류 예제를 보인다. 학습이 충분히 이루어지지 못한 경우에 주로 발생하는 오류로 파악되나 왜 반복오류가 /NG/ 음소의 경우에만 발생하는지에 대해서는 그 이유를 파악하기 어려웠다.

Table 22. Examples Repetitive Errors in NODE(IPA) → CMUdict(ARPAbet)

entry	NODE	CMUdict (truth)	CMUdict (model)	description
feeling	/fi : liŋ/	/F-IY1-L-IH0-NG/	/F-IY1-L-IH0-NG-NG/	/NG/ repetition
king	/kiŋ/	/K-IH1-NG/	/K-IH1-NG-NG/	/NG/ repetition

전체 오류의 46% 이상 발생한 모음교체는 NODE의 음소 /ɪ/가 CMUdict의 /AH/와 /IH/로 매핑되고 음소 /ə/가 /AH/와 /ER/로 매핑되면서 발생한 경우이다. Table 23에 예제를 보인다.

Table 24는 전체 사전 변환 중, 특이 사항이 있는 변환들만 다시 정리한 것이다.

Table 23. Examples of Vowel-replacement Errors in NODE(IPA) → CMUdict(ARPAbet)

entry	NODE	CMUdict (truth)	CMUdict (model)	description
fillip	/fɪlɪp/	/F-IH1-L-AH0-P/	/F-IH1-L-IH0-P/	/AH/ vs. /IH/ replacement
bracket	/brakɪt/	/B-R-AE1-K-IH0-T/	/B-R-AE1-K-AH0-T/	/AH/ vs. /IH/
fin	/fɪn/	/F-IH1-N/	/F-IH1-N/	positive
comedy	/kɔmɪdi/	/K-AA1-M-AH0-D-IY0/	/K-AA1-M-AH0-D-IY0/	positive
millar	/mɪlə/	/M-IH1-L-ER0/	/M-IH1-L-AH0/	/AH/ vs. /ER/
cotta	/kɔtə/	/K-AA1-T-AH0/	/K-AA1-T-ER0/	/AH/ vs. /ER/
hurt	/hɜ : t/	/HH-ER1-T/	/HH-ER1-T/	positive
extra	/ɛkstrə/	/EH1-K-S-T-R-AH0/	/EH1-K-S-T-R-AH0/	positive



Table 24. Error Analysis of Pronunciations Generated from the Model Between CMUdict, NODE and NOAD

		CMUdict(ARPAbet)→NOAD(NOAD)	CMUdict(ARPAbet)→NODE(IPA)	NODE(IPA)→NOAD(NOAD)
edit-dist length	error with 1 edit-dist	85	132	129
	error with 2 edit-dist	32	48	37
	error with 3 or more	7	23	13
	<b>total number of errors</b>	<b>124</b>	<b>203</b>	<b>179</b>

edit kind				
edit kind	insertion error	4	12	35
	deletion error	9	21	32
	replacement error	59	155	74

type of error				
type of error	stress error	<b>54 (43.5%)</b>	30 (14.8%)	52 (29.1%)
	vowel replacement error	36 (29.0%)	<b>117 (57.6%)</b>	38 (21.2%)
	consonant-related error	7 (5.6%)	3	<b>42 (23.5%)</b>

## (1) CMUdict(ARPAbet) → NOAD(NOAD) 변환

가장 많은 부분을 차지한 오류가 강세 오류이다. CMUdict와 NOAD의 발음 당 평균 강세 개수는 1.26, 1.24로 각각 비슷하다. 그러나 Table 25에서 보는 바와 같이 CMUdict와 NOAD의 경우, 평가 데이터 1,000개 중 강세 개수가 서로 다른 발음이 95개 나타났다. CMUdict와 NODE의 경우에는 교집합으로 나타나는 부분은 적으나, NODE의 경우 평가 데이터의 97% 이상이 하나의 강세만을 갖고 있기 때문에 오히려 강세 오류가 상대적으로 다른 사전에 비해 낮게 나타났다.

Table 26에 CMUdict → NOAD에서 발생한 강세 오류의 일부를 나열하였다.

Table 26. Examples of Stress Errors in CMUdict(ARPAbet) → NOAD(NOAD)

entry	CMUdict	NOAD (truth)	NOAD (model)	description
photon	/F-OW1-T-AA2-N/	/ˈfɒtən/	/ˈfɒt,ən/	While CMUdict has 1 <sup>st</sup> and 2 <sup>nd</sup> stresses, NOAD has 1 <sup>st</sup> only.
prefix	/P-R-IY1-F-IH0-K-S/	/ˈprɛfiks/	/ˈprɛfiks/	While NOAD has 1 <sup>st</sup> and 2 <sup>nd</sup> stresses, CMUdict has 1 <sup>st</sup> only.
afterglow	/AE1-F-T-ER0-G-L-OW2/	/ˈaftərglɔ/	/ˈaf,tərglɔ/	NOAD generates two 2 <sup>nd</sup> stresses.

Table 25. The Number of Common Stresses According to Pair of Lexicons

number of stresses	CMUdict∩NOAD / NOAD	CMUdict∩NODE / NODE
2	232 / 276	20 / 22
1	673 / 717	715 / 971
0	0 / 7	0 / 7
incompatible	95	65

Table 27. Examples of Various Mapping from /AH/ According to Lexicons

entry	from CMUdict	to		
		NOAD	MWCD	NODE
fillip	/F-IH1-L-AH0-P/	'fɪləp	'fi-ləp	'fɪlɪp
bazaar	/B-AH0-Z-AA1-R/	bə'zɑr	bə-'zɑr	bə'zɑ :
backup	/B-AE1-K-AH2-P/	'bək,əp	'bək-,əp	'bəkAp

## (2) CMUdict(ARPAbet) → NODE(IPA) 변환

이 변환의 경우 모음교체 오류가 전체 오류의 반 이상 발생하였다. 이는 두 사전의 발음 기호에서 사용하는 음소 개수가 다르기 때문으로 판단된다. CMUdict가 사용하는 발음기호 ARPAbet이 39개의 음소(phoneme)를, NODE의 IPA가 49개의 음소를 사용한다. 음소 집합의 차이가 다양한 매핑을 만들어 낼 수 있다. Table 27의 예제를 살펴보면 CMUdict의 음소 /AH/는 NOAD와 MWCD에서는 모두 /ə/로 매핑되는데 반면, NODE에서는 세 가지 모두 다른 발음으로 매핑되는 것을 볼 수 있다. 이와 같은 다양성으로 인해 NODE의 경우 모음교체 오류가 전체의 47% 이상 발생된 것으로 판단된다.

Table 28은 CMUdict → NODE 변환에서 모음교체 오류가 발생한 예제들이다.

## (3) NODE(IPA) → NOAD(NOAD) 변환

NODE(IPA) → NOAD(NOAD) 변환에서는 자음오류가 전체의 23.5%를 차지한다. Table 3과 Table 4에서 설명한 바와 같이 영국 영어 NODE는 음소 /r/을 생략하는 특징을 갖고 있으며 미국 영어 NOAD는 /t/를 /d/로 약화시켜 발음한 경우가 다수 포함되어 있다. 이런 이유로 두 변환에서 /r/ 음소 관련 오류와 /t/, /d/ 교체 오류가 다수 발생하였다. Table 29에 예제 일부를 보인다.

Table 28. Examples of Vowel-replacement Errors in CMUdict(ARPAbet) → NODE(IPA)

entry	CMUdict	NODE (truth)	NODE (model)	description
oversleep	/OW2-V-ER0-S-L-IY1-P/	/əʊvə'sli : p/	/əʊvə'sli : p/	positive
prolactin	/P-R-OW0-L-AE1-K-T-AH0-N/	/prəʊ'laktɪn/	/prə'laktən/	/ə/ vs. /əʊ/ replacement
lettuce	/L-EH1-T-AH0-S/	/lɛtɪs/	/lɛtəs/	/ə/ vs. /ɪ/
senator	/S-EH1-N-AH0-T-ER0/	/ˈsɛnətə/	/ˈsɛntə/	/ə/ vs. /ɪ/
awake	/AH0-W-EY1-K/	/ə'weɪk/	/ə'weɪk/	positive
chorus	/K-AO1-R-AH0-S/	/ˈkɔ : rəs/	/ˈkɔrəs/	/ɒ/ vs. /ɔ : /
mothball	/M-AO1-TH-B-AO2-L/	/mʊθbɔ : l/	/mʊθbɔ : l/	positive
wank	/W-AA1-NG-K/	/wɑŋk/	/wəŋk/	/ə/ vs. /a/
chock	/CH-AA1-K/	/tʃɔk/	/tʃɔ:k/	/ɒ/ vs. /ɔ : /
convoke	/K-AA0-N-V-OW1-K/	/kən'vəʊk/	/kɒn'vəʊk/	/ɒ/ vs. /ə/
crotch	/K-R-AA1-CH/	/krɒtʃ/	/krɒtʃ/	positive

Table 29. Examples of Consonant-related Errors in NODE(IPA) → NOAD(NOAD)

entry	NODE	NOAD (truth)	NOAD (model)	description
yarn	/jɑ : n/	/jærn/	/jän/	/r/ deletion
sigma	/ˈsɪgmə/	/ˈsɪgmə/	/ˈsɪgmər/	/r/ insertion
nard	/nɑ : d/	/nærd/	/nærd/	positive
politic	/ˈpɒlɪtrɪk/	/ˈpälətɪk/	/ˈpälədik/	/t/ vs. /d/ replacement
waiter	/ˈweɪtə/	/ˈwädər/	/ˈwädər/	positive

Table 30. Accuracy and Type of Errors According to the Models

Model	Accuracy <sup>4)</sup>	Type of errors			
		stress	vowel replacement	repetitive	consonant-related
1 MWCD(MWCD) → NODE(IPA)	187 (81.3%)	28 (15.0%)	<b>88 (47.1%)</b>	5 (2.7%)	7 (3.7%)
2 MWCD(MWCD) → CMUdict(ARPAbet)	180 (82.0%)	56 (31.1%)	<b>82 (45.6%)</b>	1 (0.01%)	3 (1.7%)
3 MWCD(MWCD) → NOAD(NOAD)	122 (87.8%)	<b>41 (33.6%)</b>	30 (24.6%)	0	11 (9.0%)
4 CMUdict(ARPAbet) → NODE(IPA)	203 (79.7%)	30 (14.8%)	<b>117 (57.6%)</b>	2 (1.0%)	3 (1.5%)
5 CMUdict(ARPAbet) → MWCD(MWCD)	111 (88.9%)	29 (26.1%)	<b>46 (41.4%)</b>	0	4 (3.6%)
6 CMUdict(ARPAbet) → NOAD(NOAD)	124 (87.6%)	<b>54 (43.5%)</b>	36 (29.0%)	0	7 (5.6%)
7 NODE(IPA) → NOAD(NOAD)	179 (82.1%)	<b>52 (29.1%)</b>	38 (21.2%)	1 (0.5%)	<b>42 (23.5%)</b>
8 NODE(IPA) → CMUdict(ARPAbet)	255 (74.5%)	48 (18.8%)	<b>118 (46.3%)</b>	<b>14 (5.5%)</b>	8 (2.0%)
9 NODE(IPA) → MWCD(MWCD)	162 (83.8%)	23 (14.2%)	41 (25.3%)	2 (1.2%)	<b>48 (29.6%)</b>
10 NOAD(NOAD) → NODE(IPA)	221 (77.9%)	43 (19.5%)	<b>95 (43.0%)</b>	0	29 (13.1%)
11 NOAD(NOAD) → MWCD(MWCD)	<b>104 (89.6%)</b>	19 (18.3%)	<b>37 (35.6%)</b>	0	<b>30 (28.8%)</b>
12 NOAD(NOAD) → CMUdict(ARPAbet)	185 (81.5%)	50 (27.0%)	<b>70 (37.8%)</b>	2 (1.1%)	22 (11.9%)

4) 평가데이터 1,000개 중 발음 변환이 틀린 것의 개수와 정확도이다.

Table 30은 전체 변환에 대한 정확도와 오류별 종류를 정리한 것이다.

본 실험에서 오류 분석을 통해 파악한 발음 자동 변환 seq2seq 모델은 다음과 같은 특성을 갖고 있다.

- 데이터 유사성이 높은 시퀀스 사이의 변환 모델의 성능이 우수했다.
- 반복 오류는 변환 모델의 성능이 낮은 경우에 발생할 수 있다.
- 일반적으로 길이가 짧은 시퀀스에서 길이가 긴 시퀀스를 생성해 내는 변환 모델의 성능이 우수했다.
- seq2seq model에서 버킷 개수를 늘린다고 정확도가 계속 향상되는 것은 아니다.

### 5. 결 론

본 연구에서는 seq2seq model을 이용하여 동일한 단어에 대하여 사전 간 발음기호를 자동으로 변환하는 모듈을 구현하였다.

각 사전마다 채택하고 있는 음성기호 시스템이 다르며, 동일한 단어에 대한 발음 정보 역시 다른 경우가 많았다. 가장 대표적으로 영국 영어와 미국 영어를 다루고 있는 사전은 발음 정보에 상당한 차이를 보였다. 발음의 강세 정보를 표현하는 방식에도 차이가 있었으며 강세 정보를 얼마나 상세히 다루었는지에도 차이를 보였다. 또한 사전의 개발 목적에 따라 다루고 있는 발음의 다양성 정도가 달라서 각 사전이 포함하고 있는 발음의 이음(allophone) 개수에도 많은 차이를 보였다. 사전에서 발음기호를 기술하는 방식에도 많은 차이를 보였기 때문에 적절한 전처리 과정을 거쳐 발음기호를 정규화 하는 작업이 필요했다.

Seq2seq model을 사용한 대표적인 응용 분야는 기계번역이다. 기계번역의 경우 다루는 대상 문장이나 어휘 수가 워낙 크기 때문에 그 결과에 대한 정확한 분석이 사실상 어렵다. 본 연구가 다루는 발음 정보의 경우 입력이 3~16의 범위를 갖는 다소 짧은 입력을 다루고 있으며 정확도 평가가 매우 명확하기 때문에 모델의 성능 비교를 쉽게 해볼 수 있다. 발음 정보의 자동 변환의 결과와 오류 분석을 통해 간접적으로 seq2seq model의 특성을 파악해 볼 수 있었다.

### References

[1] Jurafsky, Dan. *Speech & language processing*, Pearson Education India, 2000.  
 [2] N. M. Hosseinzadeh, A. K. Z. Kambuziyya, and M. Shariati, "British and American phonetic varieties," *Journal of Language Teaching and Research*, Vol.6, No.3, pp.647-655. 2015.

[3] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078. 2014.  
 [4] RNNs in Tensorflow, a Practical Guide and Undocumented Features [Internet], <http://www.wildml.com/2016/08/rnns-in-tensorflow-a-practical-guide-and-undocumented-features/>.  
 [5] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *NIPS-2014*, pp.3104-3112, 2014.  
 [6] A. Finch, and E. Sumita, "Phrase-based machine transliteration," in *Proceedings of the Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST)*, pp.13-18, 2008.  
 [7] A. Finch, P. Dixon, and E. Sumita, "Rescoring a phrase-based machine transliteration system with recurrent neural network language models," in *Proceedings of the 4th Named Entity Workshop*, Association for Computational Linguistics, pp. 47-51, 2012.  
 [8] A. Finch, L. Liu, X. Wang, and E. Sumita, "Neural network transduction models in transliteration generation," in *Proceedings of NEWS 2015 The Fifth Named Entities Workshop*, p.61, 2015.



### 이 공 주

e-mail : kjoolee@cnu.ac.kr  
 1992년 서강대학교 전자계산학과(학사)  
 1994년 한국과학기술원 전산학과 (공학석사)  
 1998년 한국과학기술원 전산학과 (공학박사)

1998년~2003년 한국마이크로소프트(유) 연구원  
 2003년 이화여자대학교 컴퓨터학과 대우전임강사  
 2004년 경인여자대학 전산정보과 전임강사  
 2005년~현 재 충남대학교 전자정보통신공학과 교수  
 관심분야: 자연언어처리, 기계번역, 정보검색, 정보추출



### 최 용 석

e-mail : yongseok.choi.92@gmail.com  
 2016년 충남대학교 정보통신공학과(학사)  
 2016년~현 재 충남대학교 전자전파 정보통신공학과 석사과정  
 관심분야: 자연언어처리, 정보검색, 기계학습, 인공지능

### 부 록

다음 표는 사이트<sup>5)</sup>에서 정리해 놓은 각 발음 음소별 매핑 중, 본 연구에서 사용하는 발음기호 부분만을 발췌한 것이다. 이 매핑은 완전한 것은 아니며, 가장 일반적이며 매핑을 기술한 것이다.

ARPAbet	IPA	NOAD	MWCD
ch	tʃ	CH	ch
hh	h	h	h
	hw	(h)w	hw
jh	dʒ	j	j
	x	KH	<u>k</u>
g	g	g	g
ng	ŋ	NG	ŋ
sh	ʃ	SH	sh
th	θ	TH	th
dh	ð	<u>TH</u>	<u>th</u>
y	j	y	y
zh	ʒ	ZH	zh

The following letters have the same values in all systems listed: b, d, f, g, k, l, m, n, p, r, s, t, v, w, z.

ARPAbet	IPA	NOAD	MWCD
æ	æ	a	a
e(y)	eɪ	ā	ā
eh r	ɛər	e(ə)r	ar
aa	ɑ :	ä	ä, á
aa r	ɑr	är	är
eh	ɛ	e	e
iy	i :	ē	ē
ih r	iər	i(ə)r	ir
ih	ɪ	i	i
ay	aɪ	ī	ī
aa	ɑ	ä	ä
ow	oʊ	ō	ō
ao	ɑ :	ô	ô
	ɔr	ôr	ôr
	ɔər		
oy	ɔɪ	oi	ói
uh	ʊ	ōō	ú
uh r	ʊər	ōōr	úr
uw	u :	ōō	ü
aw	aʊ	ou	aú
ah	ʌ	ə	ə
er	ɜr, ɝ :	ər	ər
ah	ə	ə	ə
ah	i	ə, i	ə
er	ər	ər	ər
y uw	ju :	yōō	yü
	a		á
	ø : , œ	œ	œ, œ
	y : , Y	Y	ÿ, ue
	ʒ	ôN	ô <sup>n</sup>

5) [https://en.wikipedia.org/wiki/Pronunciation\\_respelling\\_for\\_English](https://en.wikipedia.org/wiki/Pronunciation_respelling_for_English)